

Biosensors and Bioinformatic for Genetic Analysis

Recent developments in genetics and biochemistry, propelled by large-scale initiatives like the Human Genome Project, have led to new approaches to the analysis of interactions between bio-molecules (DNA, proteins) that are assuming a strategic role in enabling advances in medical, biological, environmental applications. In particular, thanks to the evolution of micro-fabrication technologies (derived from the rapidly-developing microelectronics technology), it is now possible to design and fabricate small-scale devices that can perform a large number of biochemical tests in parallel. These techniques, coupled with high-throughput sensing, signal processing devices, and atomic force microscopes can produce huge amounts of data in a matter of hours. The management of this amount of information determines the increasing development of data mining technologies to automate to some degree, biological and medical information discovery from experiments and publicly accessible databases.

In this context, there is a great demand of cheap, portable, and easy to use devices for point-of-care analysis. Micro-fabricated systems with integrated-sensor techniques can satisfy these requirements. These systems integrate on the same substrate the sites of the reaction, the sensors providing electric signals and the circuits for signal conditioning and amplification.

The **Bio** research area of Micrel Lab group aims at developing **Electronic Systems** and **Data Mining Techniques** for genetic analysis and information discovery.

Point-of-care Electronic Systems for Genetic Analysis

The first generation of micro-arrays, based on fluorescence detection techniques finds now wide usage in fundamental research in genomics and proteomics. The diffusion of micro-arrays in medical, industrial, environmental applications is however currently limited by high costs, and by usage and data analysis complexity. This research project targets the design and development of innovative integrated devices that overcome the limitation of current micro-array, in particular avoiding the insertion in the material to be analysed required of the markers required for optical detection, producing increased cost of the devices and more complex processes for sample preparation. To this purpose, we will study two techniques for detection of DNA hybridization with complementary aspects as far as realization, use and cost. These techniques aim at detecting the differences induced by DNA hybridization on the capacitance of a two electrode structure, particularly suitable to be microfabricated (a) and the absorption of UV radiation by a layer of DNA deposited on top of semiconductor UV sensor (b), respectively. In both cases, the biochemical reaction results in an electrical signal that needs to be detected, conditioned, converted and processed in digital form. To this purpose, our research team study, design and realize all the electronics that is needed, with the aim to minimise cost, guarantee performance and reliability and simplify the use of the final system. Finally, we will design and develop active (smart) sensor architectures with high parallelism and miniaturization, which can perform a large number of biochemical analyses in parallel on a single microfabricated substrate, with clear cost and throughput advantages. The projects are characterized by strong multi-disciplinary because they requires competences in electronics, physics, biochemistry, microfabrication technology. The Project has a strong multi-disciplinary character, in that it requires qualified know-how in electron devices (sensors), electronic circuits, technology and biochemistry. Micrel Lab has a long experience in the first three of these fields, while for the latter they exploits the competences of the group of biochemistry active at the University of Bologna, with which a fruitful cooperation on this subject has started more than two year ago.

C. Guiducci, C. Stagni, G. Zuccheri, A. Bogliolo, L. Benini B. Samorì. B. Riccò, *A Biosensor for Direct Detection of DNA Sequences*, Proc. of ESSDERC2002

C. Guiducci, C. Stagni, G. Zuccheri, A. Bogliolo, L. Benini B. Samorì. B. Riccò, *DNA Detection by integrable electronics*, Biosensors and Bioelectronics 19 (2004) 781–787

C. Stagni, C. Guiducci, G. Zuccheri, L. Benini. B. Riccò, *Fully electronic DNA detection technique*, Proc. of AISEM 2004

C. Guiducci, V. Stambouli-Sene, M. Labeau, L. Benini, B. Riccò, *Conductive oxides as new materials for electrical DNA detection*, Proc. of Biosensors 2004

Data Mining Techniques - Computational Biology

The projects in the bioinformatic area aim at devising automated analysis procedures for significant biology problems as the development of techniques for the extraction of DNA structural properties through high resolution microscopy, as the studies of clustering and partitioning algorithms from microarray experiments, as the discovery of genetic networks for the analysis of the global behaviour of cellular networks and the generation of relative predictive models. For these purposes powerful combinatorial analysis and optimization techniques developed in the design automation domain, neural networks and image processing techniques have to be employed. We are working on three main projects:

- Development of automated techniques for DNA structural feature analysis in Atomic Force Microscope images
- Microarray clustering
- RNA silencing and Gene Regulatory Networks design

Techniques for automated analysis of DNA molecules in AFM images

The Atomic Force Microscopy is characterized by high resolution and high signal-to-noise ratio, so it can be applied to nucleic acids. Furthermore it allows direct visualization of single DNA molecule without contrast-enhancing agents and thus without altering the structure of the molecules. For these reasons it is very effective and useful to studies the characteristics and the properties of the DNA molecules.

We developed automated algorithms for DNA molecules feature analysis and extraction through a set of fully automated image processing steps, pattern recognition techniques and noise analysis and deletion procedures. Moreover, we employed ad-hoc combinatorial optimization techniques to investigate and reconstruct DNA molecule structural properties such as the intrinsic curvature and the flexibility. Our goals using these automated algorithms are the identification of a specific DNA target, the construction of physical genome maps and genotyping, the investigation of DNA molecule structural properties, the analysis of DNA secondary structure transitions and the DNA-protein interactions, and the investigation of transcription rules.

Microarray Clustering and Cluster Biological Evaluation of Gene Expression Data

The use of microarray is now becoming more and more widespread among researchers in biology, medicine and computer science, for two main reasons.

First, from a biomedical point of view, their important flexibility allows scientists to address with this tool a wide variety of biochemical-medical questions: from the individuation of subtypes of diseases, to the understanding of biochemical processes in both healthy and perturbed states.

Second, from an algorithmic point of view, the high throughput of the information obtained has required new efforts in the data mining and in particular in the process of clustering genes. The classical algorithms (hierarchical cluster, k-means etc.) are not able to capture the coherent behavior on subset of dimension (experiment). Several algorithm have been proposed. Our enhanced pCluster algorithm allows the mining of high quality cluster, that is clusters highly coherent and variable (non-constant) in their expression level profile. Observing the clusters through the categories of Gene Ontology we could prove that our clusters present a coherent biological meaning (purity).

However, besides the clustering step, other issues need to be addressed.

Microarrays are now widely used in *Translational medicine*: this is a field of medicine in which knowledge and techniques from different fields (physics, mathematics, medicine, biology, computer science) are integrated and interconnected in a clinical focused manner to impact diagnosis and treatment of diseases. Genomic information retrieved from microarrays are introduced because of the very precise description it can provides of the interested disease, thus allowing for example the matching between non-invasive clinical parameters and precise genetic pathways. The correlation of these information requires the integration and automation of new algorithm and statistic approaches. We are currently working on this approach.

References:

- *Biclustering of Expression Data*, Y. Cheng, G. M. Church, ISMB'00, 2000
- *Clustering by Pattern Similarity in Large Data Sets*, H. Wang et Al, SIGMOD 2002
- *DNA Arrays for Analysis of Gene Expression*, P.O.Brown, M.B. Eisen, <http://cmgm.stanford.edu/pbrown/>
- *Endothelial cell Diversity revealed by global expression profiling*, Chi, Chang, Haraldsen, PNAS, September 16 2003, vol100, n.19, 10623-10628
- *Gene expression pattern of Breast Carcinomas distinguish tumor subclasses with clinical implication*, Sorlie et Al, , September 11 2001, vol98, n.19, 10869-10874
- *Diversity of gene expression in adenocarcinoma of the lung*, Garber et Al. PNAS November 20, 2001 vol 98 n.24 13784-13789

RNA silencing and Gene Regulatory Networks Design

Biology is currently undergoing a shift from a mostly qualitative to an information intensive, quantitative science. Using large-scale biological analysis technologies, we are gaining global views of structural and dynamic information in the form of whole genome sequences and the corresponding gene activity and expression patterns at the RNA and protein level.

These data reflect the workings of a complex molecular information processing system. In many ways these systems can be effectively viewed from the perspective of genetic feedback networks, given that the fundamental step of biological information flow resides in gene activation and its control through the activity of regulatory genes. One of the greatest challenges in modern biology lies in interpreting a large mass of analytical data to discover genetic networks and generate predictive models.

One of the most effective tools to obtain Gene activation information are DNA Microarrays, which provide a global insight into the transcriptional behaviour of the organism. The deciphering of the regulatory mechanism based on the transcript profiles has opened the door for the study of global mechanisms in the cell. By measuring thousands of genes simultaneously it has become possible to get snapshots of the states of a cell. While monitoring the mRNA levels reveals only a part of the whole picture and protein arrays are still in their infancy, the DNA microarray technique has quickly become an established method and will lead the way in the analysis of the global behavior of cellular networks.

In this context, we are currently working on

- *RNA silencing molecule design*: design of models of small RNA-specific molecules for RNA interference to inhibit the expression of relative transcripts. This allows inferences for the

regulatory networks for systematic analysis of gene expression and function, and for therapeutic gene silencing and drug design.

- *Gene Regulatory Networks inference and discovery*: we use microarrays to derive a dynamic information about gene activities, that enables to move beyond simple model of the putative interactions of small sets of molecules. The objective is to define specific biochemical pathways and to obtain systematic, quantitative approaches to functional inference and modeling. Theoretical investigations into genetic network behavior require computational tools for network construction, visualization and simulation.

E. Ficarra, L. Benini, E.Macii, G. Zuccheri, *A Robust Algorithm for Automated Analysis of DNA Molecules in AFM Images*, Proc. of IASTED Biomedical Engineering (BioMED 2004)

D. Masotti, E. Ficarra, L. Benini, E.Macii, *Techniques for Enhancing Computation of DNA Curvature Molecules*, Proc. of IEEE Fourth Symposium on Bioinformatics and Bioengineering (BIBE2004)

E Ficarra, L Benini, B Riccò, G Zuccheri, *Automated Dna Sizing in Atomic Force Microscope images*, Proc. of IEEE International Symposium on Biomedical Imaging (ISBI 2002)

E Ficarra, D Masotti, L Benini, M Milano, A Bergia, *Automated Dna curvature profile reconstruction in Atomic Force Microscope images*, AI*IA Notizie, Vol. 4, n. Dec., pp. 64-68 (2002);

E. Ficarra, L. Benini, E.Macii, G. Zuccheri, *Automated DNA Fragments Recognition and Sizing through AFM Image Processing*, submitted for publication on IEEE Transactions on Information Technology in Biomedicine;

E. Ficarra, D. Masotti, L. Benini, E.Macii, G. Zuccheri, B. Samori, submitted for publication on IEEE Transactions on Biomedical Engineering

S. Yoon, C. Nardini, L. Benini, G. de Micheli, *Enhanced pClustering and Its Applications to Gene Expression Data*, Proc. of IEEE Fourth Symposium on Bioinformatics and Bioengineering (BIBE2004)